Cataloging Internet resources*

By Melinda Reagor Flannery, M.Div., M.L.I.S.
Head of Cataloging

Rice University
Fondren Library MS 44
6100 Main Street
Houston, Texas 77005-1892

The number of resources available on the Internet continues to expand exponentially, but finding appropriate resources is still a fragmented, hit-or-miss operation. Traditional library expertise in bibliographic description and access should be applied to the management of this emerging body of material. In the process, catalogers will be able to assess the adequacy of current tools (e.g., cataloging codes, machine-readable cataloging formats, integrated library systems) for providing access to Internet resources and will contribute credibly to design or redesign of access tools. This paper outlines the major issues that must be considered in cataloging electronic resources.

INTRODUCTION

Exciting developments are taking place in many areas of the new world of Internet resources, and it is clear that many useful things can be done with a given Internet resource other than catalog it. And yet, libraries traditionally have relied on cataloging to describe and analyze for retrieval the majority of materials deemed worthy of collection and preservation. The role of cataloging in the new electronic environment is as yet undetermined. This paper outlines the major issues of electronic-resource cataloging and provides an account of the discussion, experimentation, and debate taking place in cataloging circles. In conclusion, a few suggestions for future action will be offered.

NATURE OF INTERNET RESOURCES

In examining cataloging issues, the initial focus must fall on the electronic resources themselves and the current Internet environment, in order to delineate how they are different from and similar to the print resources for which technical services traditions have been shaped. Those who have been exposed to the Internet know that it includes a broad variety of resources. Some may be produced locally, most are remote; some are archived, more are displaced routinely by newer versions; some are stable and even backed by long-term institutional commitment, others are here today and gone tomorrow. Many local Gophers offer very little in the way of original resources, instead offering access to resources created by others elsewhere. Finally, even those of us not responsible for collection development can see that the permanent, archival value of many—even most—Internet resources is questionable. The experience is a bit like getting the key to someone else's utility shed.

In some ways, the familiar world of print shares some of these characteristics. For example, a stack of materials on a coffee table might include a restaurant flyer, the operator's manual for a new blender, a parent's love letters, a local newspaper, the works of Shakespeare, and a Bible. The Internet has begun to provide an even broader and more eclectic range of material in immediate and worldwide distribution, with far fewer clear cues about content and quality.

THE ELECTRONIC LIBRARY MISSION

Even before addressing the cataloging question, it is worth asking what the library mission is in such an environment. It is still true that if individuals could obtain all the materials they wanted and needed—the "good stuff"—anytime they wanted without cooperating or sharing, there would be no need for libraries. Further, if they could find this good stuff reliably when they needed it, there would be no need for cataloging.

Broadly speaking, neither the library mission nor the narrower cataloging mission has been widely acknowledged yet in the electronic environment. The new search tools are just good enough and users are

* Presented at the Ninety-Fourth Annual Meeting of the Medical Library Association, San Antonio, Texas, May 16, 1994.

able to find just enough wanted material twice so that more durable solutions are not yet being demanded. An additional factor is current legislative discussion, which may lead to significant change in the cost model of Internet operation; such change surely would affect both what material is created and what is preserved. Yet librarians, catalogers among them, can evaluate the adequacy of the existing tools and see how fast the electronic "piles" are growing, and they can predict a possible crisis of access—a crisis towards which pioneering efforts and noble experiments should be directed now.

THE PROFESSIONAL CHALLENGE

Librarians, it seems, have not played the lead role in developing any of the tools currently used in navigating the Internet. The tools are various and all relatively new; their long-term adequacy as aids for navigating the volatile, expanding body of Internet resources already is questioned by creators and users alike. Wide-area information servers (WAIS) allow full-text searching of databases; Archie allows searching of anonymous file-transfer protocol (FTP) archives; and Gopher facilitates access to a broad range of primarily textual electronic resources through customized, locally developed menus. Veronica permits searching of all Gopher menus at a given point in their fluid existence; and Mosaic supports access through a graphical user interface to the broad range of information—including encoded text, images, and sound—in the World Wide Web (WWW). Standard generalized mark-up language and the simplified hypertext mark-up language support mark up of electronic texts for precise retrieval and analysis of textual segments, such as pages and paragraphs, as well as specific linguistic elements.

Although many librarians are becoming acquainted with these Internet navigation tools and are using them to create local tools, such as local Gopher servers and WWW home pages, there is a simple reason librarians have not taken the lead in innovation to meet these new challenges. Most of the development work to date has been done in computing centers, and librarians have not been in the room. Librarians are just beginning to forge some of the alliances and share some of the expertise that may help build improved tools. They are starting to find opportunities to share with computing professionals the basic principles of library service: collection development and preservation, description and access, and classification and subject vocabulary.

There is another reason librarians have not been first to collect and organize Internet resources. Librarians do not come from a professional culture that encourages wading into chaotic situations, especially those that do not seem to fit known solutions precisely. Looking at the Internet, one can find many reasons not to catalog. For example, it is unclear who is responsible for storing resources and making sure future generations can read them. There are few clear provisions for ensuring that resources undergo no unwanted changes, deliberate or inadvertent. Similarly, there is a lack of clear publishing conventions to parallel those that have been basic signposts in print collection development.

Only in large, stable electronic services, such as CANCERNET and the Human Genome Data Base, and in some commercial products is there the kind of long-term institutional commitment that has been associated with and expected of libraries. Other electronic services, Gopher servers, and electronic documents exist and pass away without reference to the reader or the library and without broad agreement among sites as to cooperative aims and responsibilities. Librarians have not been able to imagine collecting and organizing materials in such an environment, and most have not done so.

Librarians are in a position not so different from that faced by their forebears when libraries were new. And the profession has a responsibility, given that it already has taken on the preservation and management of the human record to apply known solutions to this new phase of the human record, to see if and how the old cataloging terms and solutions work.

NATIONAL DEVELOPMENTS

Recently, the OCLC Online Computer Library Center's Office of Research received a grant from the U.S. Department of Education to study the nature of Internet resources. In the spring of 1992, representatives from that project; from the Committee on the Machine-Readable Form of Bibliographic Information (MARBI)—the group responsible for preserving and modifying the U.S. machine-readable cataloging (USMARC) formats; from the Library of Congress; and from the Online Audiovisual Catalogers conducted an experiment. They used the current cataloging rules, Anglo-American Cataloging Rules, second edition, revised; and the existing USMARC format for computer files to describe the new types of electronic files appearing on the Internet. The MARBI group already was working on a proposal for format changes to accommodate the new resources, recognizing that their defining characteristic was remoteness: they were not physical and could not be held by libraries, stored in the stacks, or issued to patrons in the usual ways. In the experiment, OCLC issued to thirty volunteers samples of Internet resources to catalog using existing tools; the volunteers were to record their problems and questions. Each file was cataloged by at least three catalogers, and the results were compared [1].

The OCLC research project resulted in proposed changes to both the cataloging rules and the USMARC formats. Changes recommended in descriptive cataloging included adding appropriate terminology to the computer-files format, which was designed for large data sets and only subsequently expanded to reflect the now-more-prevalent software packages. A more far-reaching recommendation was to add a tag to the MARC-holdings format to accommodate "electronic location and access" information.

It is important to remember that cataloging-rule changes must be funneled through a set of bureaucratic channels; namely, the Cataloging Committee on Description and Access (CC:DA), then through MARBI and the bibliographic utilities. Simple format changes can be initiated through MARBI but then must be supported by the utilities and the integrated-library-system vendors to be considered fully implemented.

As of May 1994, the recommended descriptive changes were still under review by CC:DA, in the context of a broader review of the computer-files format. MARBI was able to approve the new location and access tag and has made important additions such as subfields to facilitate immediate access to digital text and image files from the USMARC records that describe them. One of these subfields accommodates the Uniform Resource Locator (URL), an element developed by the Internet Engineering Task Force (IETF) and already in heavy use as a precision finding tool used with Mosaic.

The latest USMARC update containing these changes is due in the summer of 1994, after which time the tag should be approved for use on the utilities and may find more expanded support on local systems. A more stable element, Universal Resource Name (URN), is still under development by the IETF. More suitable than is URL for permanent storage in a bibliographic record, the URN would require some manipulation and linkage with location information at the time of a search, probably through computer software not yet developed.

There are, no doubt, reasons other than the usual slow speed of bureaucracy for these changes in technical services infrastructure to be taking root so slowly. Monitors of the list AUTOCAT know that many subscribers have expressed the opinion that cataloging of Internet resources is misguided. The experience of Rice University in cataloging Internet resources provides further evidence that such experiments may provoke controversy and challenge old and new alliances within the library and beyond.

THE RICE UNIVERSITY EXPERIMENT

In the winter of 1993, Beth Shapiro, director of the Fondren Library at Rice University, began to bring up the idea of cataloging Internet resources. After some general stalling, in good academic fashion, a committee was set up to study the issue.

Two significant concerns were apparent. One was the fact that experimental cataloging is time consuming, and the need to process traditional materials was not diminishing. There was concern that if the experiment proved successful, a new and immense backlog of electronic resources requiring cataloging might develop. The other widespread concern was related to the instability of the electronic environment. The catalog always had been the permanent record of what the library offered to patrons—materials purchased and marked and physically available on the shelves. The creation of bibliographic records was not clearly appropriate for materials that moved and disappeared at such alarming rates. There was no indication that Rice University was considering making a long-term commitment to archiving any electronic resource, such as an electronic journal, judged to have lasting value. There was thus no firm guarantee that resources held and maintained remotely, once catalogued, would continue to be available to users of the Rice online catalog.

The developer of the Rice campus Gopher, Prentiss Riddle, and this author had agreed privately that electronic resources were simply too volatile to be enshrined in the online catalog. Dr. Shapiro's answer to an informal recommendation came back: the cataloging experiment was to proceed, regardless. There were certainly relatively stable resources, such as the campus directory and local policy documents, on Riceinfo, the Gopher server. Project Gutenberg was a stable, widely known project that had been in existence for many years; perhaps some of its files would be appropriate. Whatever materials were chosen, electronic resources were to be cataloged. The instability of the electronic environment could not be used as a reason or excuse for doing nothing. If the experiment proved misguided, it could be reversed.

So a group of catalogers painfully cataloged twelve Internet resources, including two documents locally produced at Rice and ten Project Gutenberg files. The coordinator of collection development personally chose all twelve files, the other bibliographers being lukewarm or even actively opposed to the idea. Policy support for the effort was coordinated by an Electronic Resources Cataloging Advisory Group, composed of representatives from collection development, reference, technical services, and cataloging.

Three catalogers and the head of cataloging constituted the Electronic Resources Cataloging Implementation Subgroup. After the initial selection of resources, each member was assigned several resources. They were to draft records and share them with the group, along with any questions or broader issues raised by the assignment. Many questions emerged

as the experiment proceeded. An early problem was deciding whether to consider a given resource published or unpublished. It was relatively easy to decide that Project Gutenberg files and the *Rice Campus Directory*, including the telephone directory, were published, but it was less clear whether the in-house document *Rice CWIS Management Proposal* should be considered published.

There was also the issue of which version of a file to catalog. Generally, the acquisitions staff orders and catalogers catalog those versions of library materials requested by the collection development staff. In the cataloging process, additional options emerged that called the original request into question. Files might exist in both compressed and uncompressed formats. A requested file might be superseded by a newer version before the cataloging record had been completed. The archive site might have a different version of the file than was available via Riceinfo, the local Gopher, which might be pointing to a source with a different or unknown schedule of updating. Electronic versions of files were found in many cases to be less current than versions available commercially, for obvious economic reasons. It was not always clear when such discoveries required group consultation and much consultation took place.

In general, the group decided to catalog the version available through Riceinfo, because that was the means through which the greatest number of patrons could view the cataloged resources. The Gopher was accessible to all library patrons, including those without electronic mail access, as a menu option on specially configured workstations in the reference room. For cataloging purposes, these workstations were referred to as LIBRIS+ (building on the name of the online catalog, LIBRIS). Riceinfo and LIBRIS+ became the two electronic locations where each cataloged resource was stored.

The use of a traditional library computer system designed for physical resources added still more challenges to the cataloging project. After much discussion with collection development and reference colleagues, call numbers were omitted from the records. The aim was to reduce the possible confusion for patrons, a consideration that outweighed the possible value of having access by classification to all resources, physical and electronic. Attempts to indicate circulation status offered a few comic moments until a way was found to generate the customized patron message, "electronic access," in the space on the brief record reserved for circulation information.

Views varied widely concerning the amount of access information that would be useful in the records, especially considering the volatile nature of the information. Because menu structure in Riceinfo might change, providing detailed menu information in a permanent cataloging record seemed questionable to some. Instructions for transferring the resource might not stay current, even if the archived version were the same as the Riceinfo version and thus properly cataloged on the same record. Some group members argued that users likely to discover electronic resources through the online catalog would not be sophisticated enough to follow detailed instructions without help, even if they remained current. Those for whom such instructions would be helpful would not look for them in LIBRIS+. The group finally decided to give a modest amount of access information in a note following an introductory phrase, "as of [date of selection]." The access note does not appear on the brief record that is the system's default display.

The group shared revised drafts of the catalog records with all library staff members to get their input, which was helpful in resolving many of the technical and philosophical problems. The first group of twelve records was produced on OCLC in November 1993.

CONCLUSIONS

In summary, this type of electronic cataloging experiment can be of value to anyone who is ready to face the momentous challenges our culture is now offering. Librarians and catalogers must think about how the human record will be documented, preserved, and accessed in the electronic age, while continuing to service the known, more traditional formats. Considering the continuing traditional workload and lack of clear guidance in any area of electronic library service, including collection development, a small experiment likely could be initiated in most libraries without growing to an unmanageable scale.

The unresolved issues become clearer, and the continuing work of national library and nonlibrary groups has deeper relevance when confronting the online-catalog message "check shelf," because the cataloged Internet resource has no bar code; when one ponders whether the file version available via the local Gopher can be cataloged on the same record as the version that can be transferred from an archive site; when one struggles to decide how little of the highly mutable access information should be put into the permanent cataloging record to help the average client, whoever that is. These questions create opportunity for librarians. For the first time in a long time, catalogers now have a chance to ask how they are doing, circulate a record and ask for comments, and argue with colleagues about the needs of clients and the integrity of the catalog.

Trying to "stuff" electronic resources into the MARC format may in the long run prove a failure, but new tools generally are not developed before old tools have been tried and found wanting. Concrete action in the form of cataloging experiments may not ma-

terially advance the important discussion of electronic access, but it will accomplish more than either naysaying or smug inattention. Whether the answer is the MARC format or some other solution, catalogers have an important part to play now in the cultural transition in which libraries are being swept along and which it is their professional responsibility both to observe critically and to shape.

ACKNOWLEDGMENTS

The author wishes to acknowledge the contributions to the experiment by fellow members of the various Rice University committees: Linda Keating, Kerry Keck, Kiem Le, Sara Lowman, and Heather Phillips.

REFERENCE

1. Assessing information on the Internet: toward providing library services for computer-mediated communication. Dublin, OH: OCLC Online Computer Library Center, 1993.

Received October 1994; accepted November 1994